

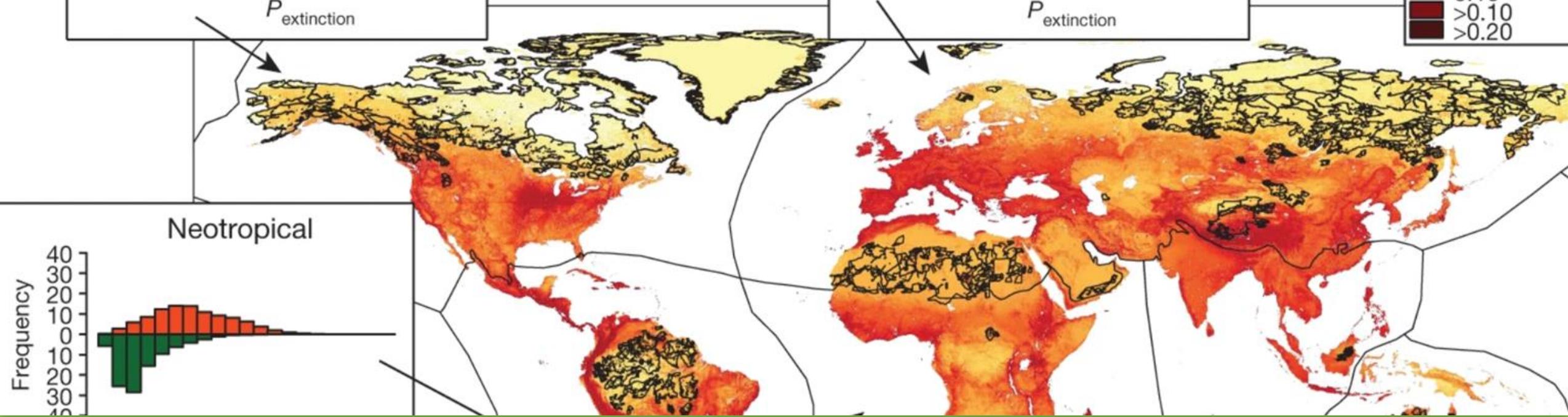
# L'utilisation des données sur la biodiversité

Andrew Rodrigues | Programme Officer



Photo: Rosette lichen (*Physcia tenella*), Isenvad, Jutland, Denmark, 6 January 2019. Photo by Lars Sørensen Grønbjerg

CC-BY-NC-FringedC via Danish Mycological Society <https://www.gbif.org/occurrence/2236587237>



## PLAN DE COURS

1<sup>ère</sup> Partie - Naviguer sur [www.gbif.org](http://www.gbif.org)

2<sup>ème</sup> Partie - Problèmes courants de qualité des données

3<sup>ème</sup> Partie - L'API

4<sup>ème</sup> - Utilisation de R

Ressources: <https://docs.gbif.org/course-data-use/en/key-documentation.html>

### Instructeurs

Andrew Rodrigues – GBIF Secretariat, Programme Office for Participation and Engagement

John Waller - GBIF Secretariat, Data Analyst

# Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF DENMARK

Occurrence records  
1,676,825,999

Datasets  
57,757

Publishing institutions  
1,665

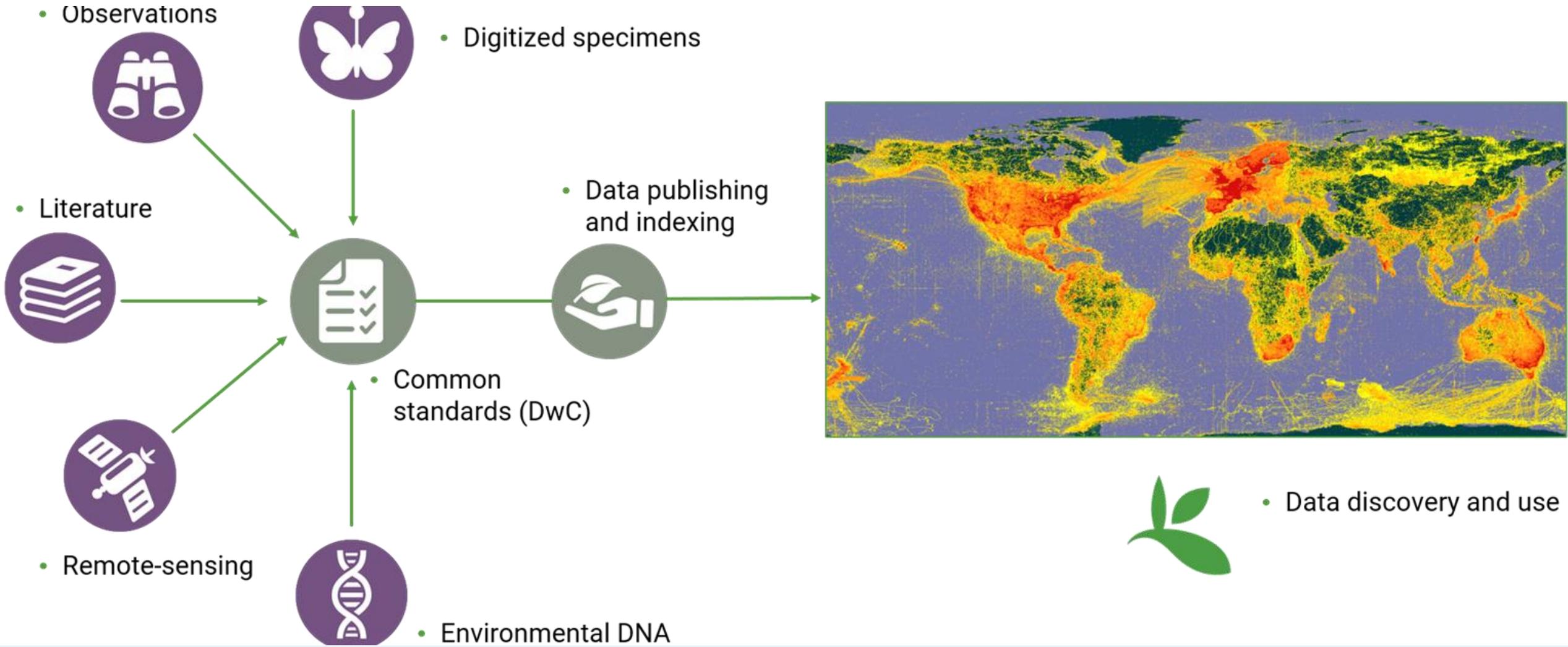
Peer-reviewed papers using data  
5,703

## POURQUOI DOIT-ON TRAITER LES DONNÉES?

Nous voulons adapter notre téléchargement de données à nos propres besoins

- Supprimez les données erronées, par ex. valeurs aberrantes
- Assurer un niveau de précision suffisant dans les données pour réaliser notre analyses

# UNE VUE FACTUELLE D'OÙ LES ESPÈCES SE SONT TROUVÉES ET QUAND



# Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF DENMARK

Occurrence records  
1,676,825,999

Datasets  
57,757

Publishing institutions  
1,665

Peer-reviewed papers using data  
5,703

## POURQUOI DOIT-ON TRAITER LES DONNÉES?

Chaque fois que vous traitez un jeu de données à utiliser, vous devrez considérer

1. Les exigences de votre analyse
2. L'équilibre entre la qualité des données et la robustesse de votre analyse

Cela peut être un processus itératif

# Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF DENMARK

Occurrence records  
1,676,825,999

Datasets  
57,757

Publishing institutions  
1,665

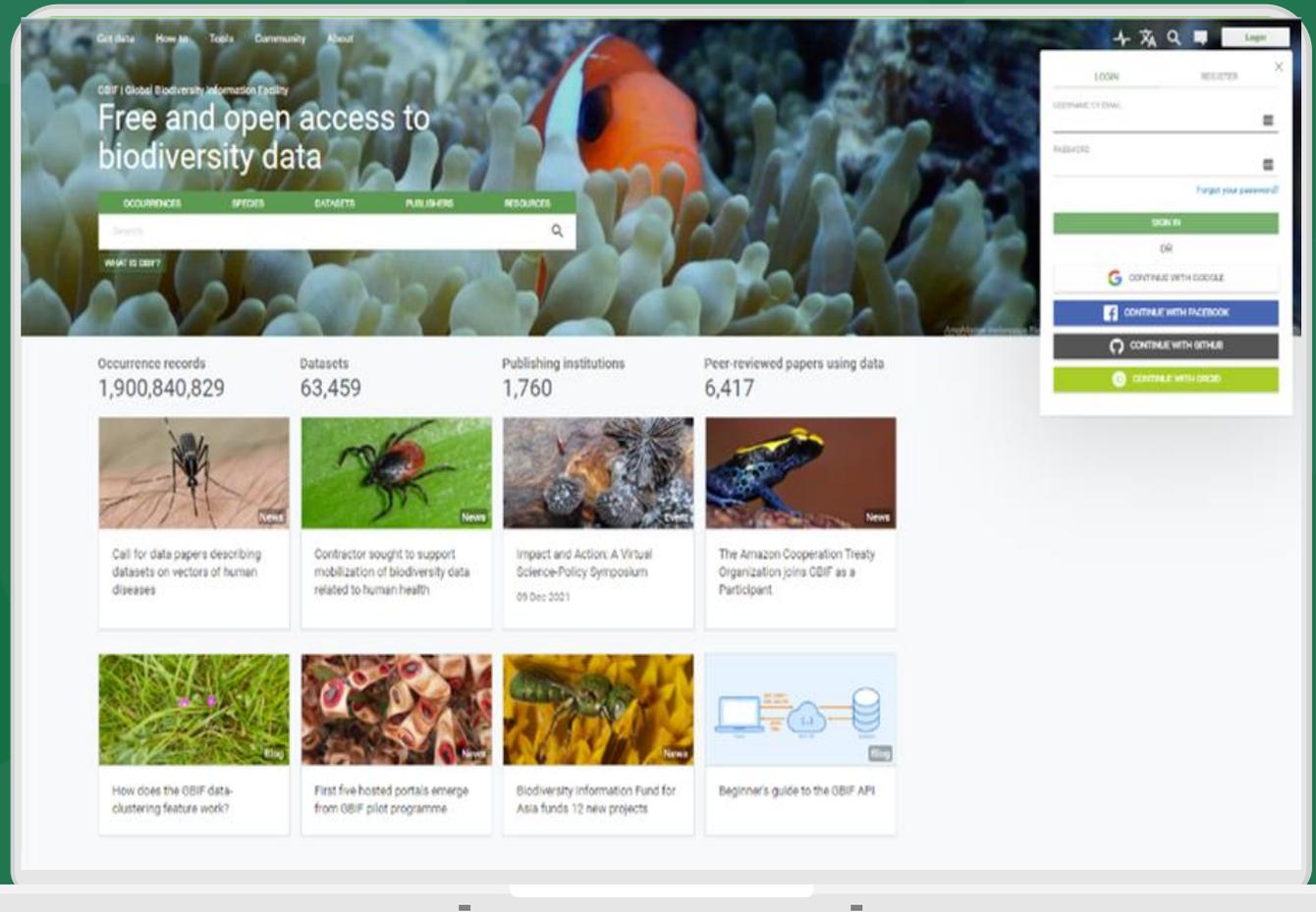
Peer-reviewed papers using data  
5,703

## RÈGLES D'OR DE L'UTILISATION DES DONNÉES ACCEDÉES VIA GBIF

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)
2. On doit accepter l'accord d'utilisation des données - <https://www.gbif.org/terms/data-user>
3. On doit documenter la façon dont vous traitez vos données
4. On doit citer correctement les données que vous utilisez
5. On doit déposer les données utilisées dans un référentiel public

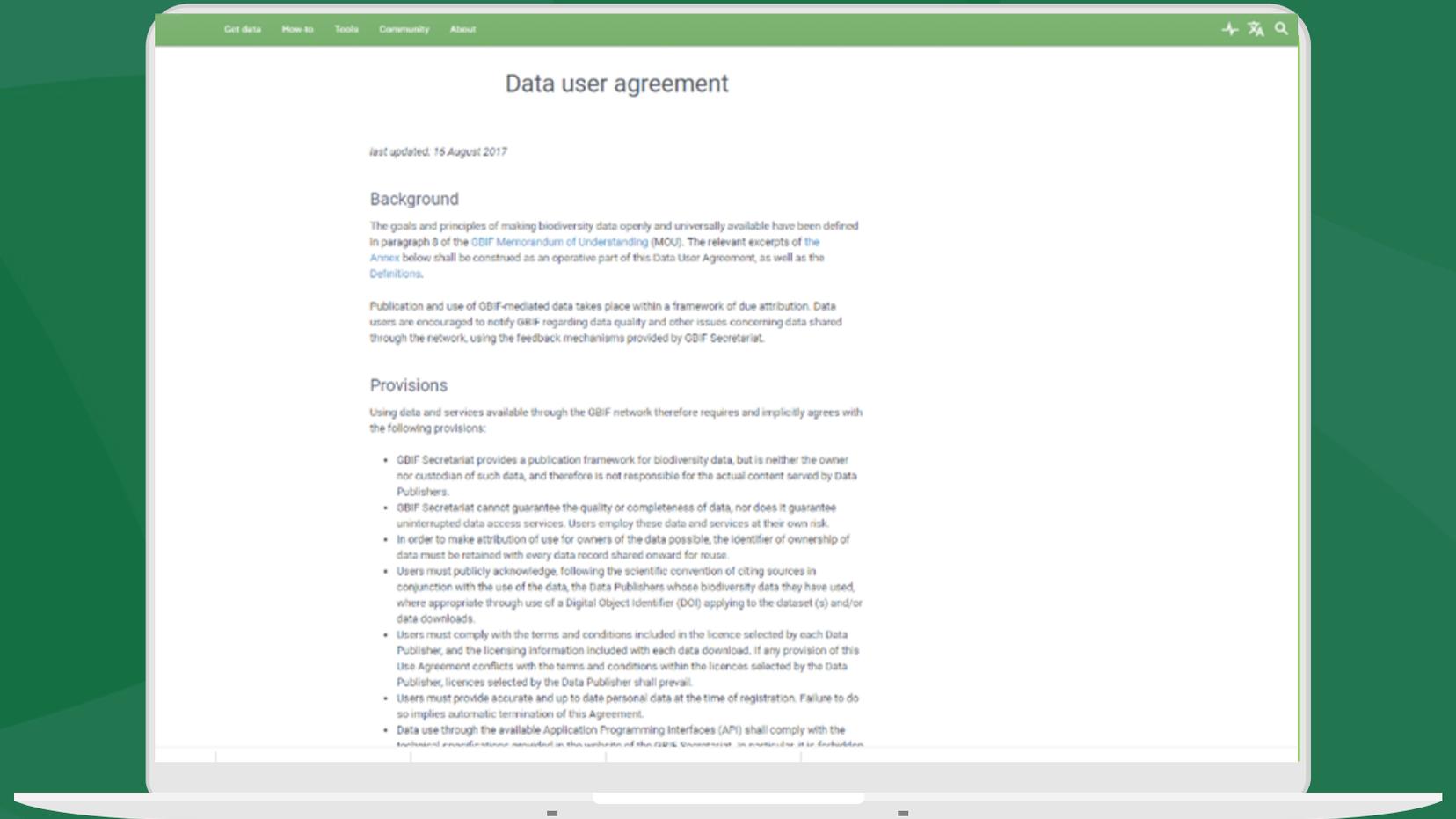
# RÈGLES D'OR DE L'UTILISATION DES DONNÉES ACCÉDÉES VIA GBIF

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)



# GOLDEN RULES OF GBIF-MEDIATED DATA USE

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)
2. On doit accepter l'accord d'utilisation des données - <https://www.gbif.org/terms/data-user>
  - Sans engagement
  - Il établit les principes directeurs de l'utilisation des données, y compris la citation des données



# RÈGLES D'OR DE L'UTILISATION DES DONNÉES ACCEDÉES VIA GBIF

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)
2. On doit accepter l'accord d'utilisation des données - <https://www.gbif.org/terms/data-user>
3. On doit documenter la façon dont vous traitez vos données

## Step 1

Download occurrence records for species x with an associated DOI

## Step 2

Remove all records from outside its native range

## Step 3

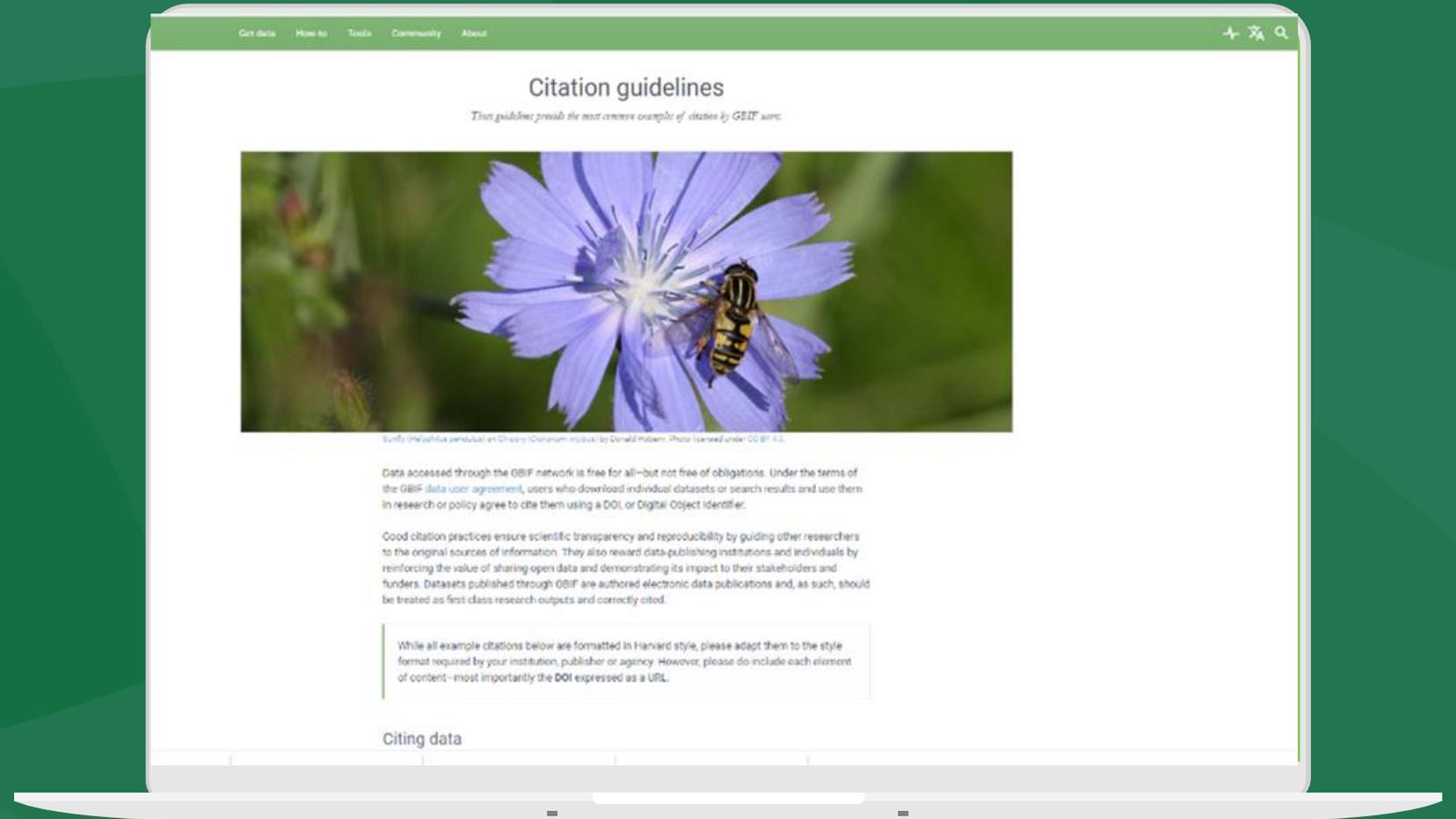
Remove all records collected before 1950

## Step 4

Final clean dataset

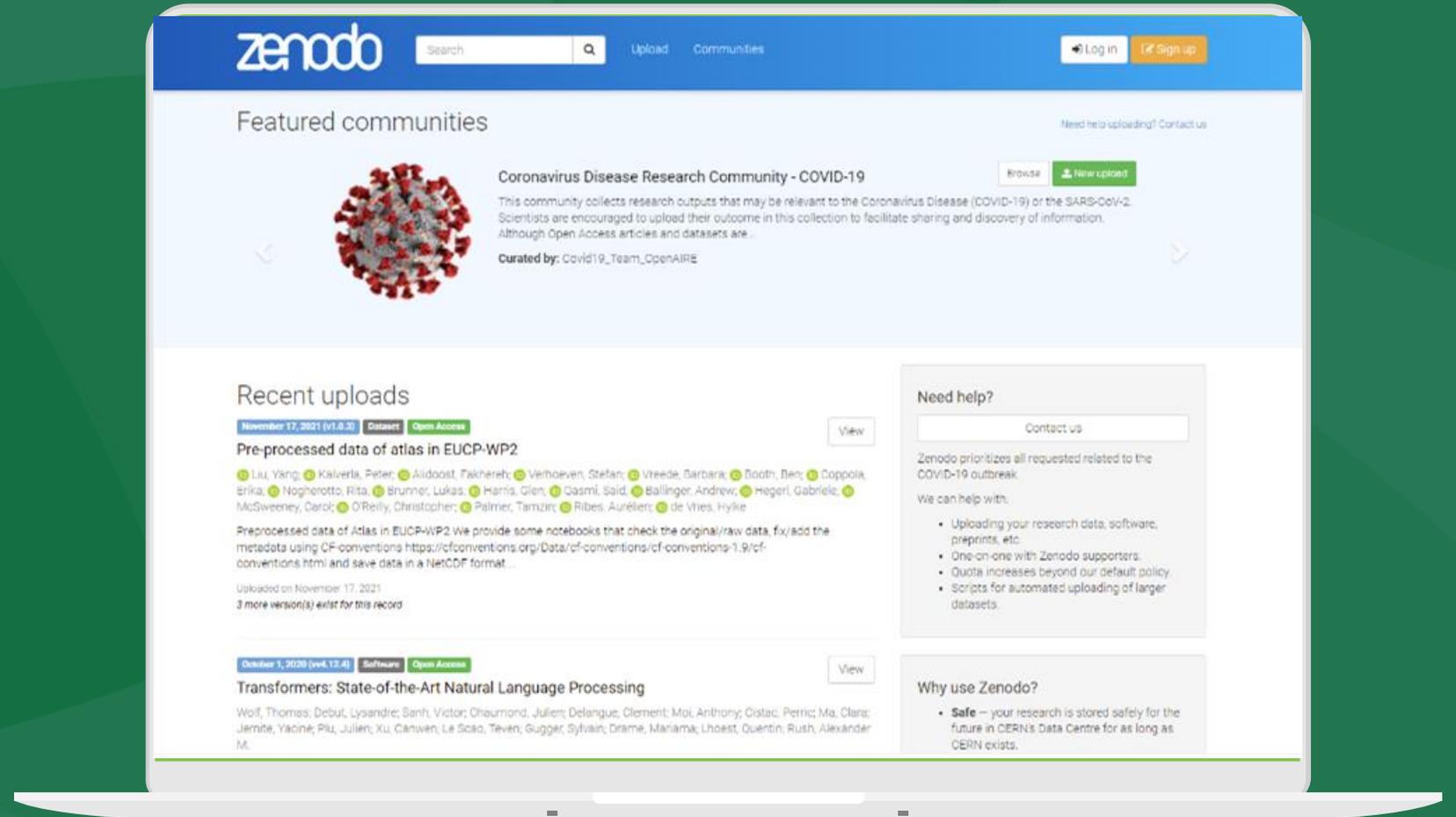
# RÈGLES D'OR DE L'UTILISATION DES DONNÉES ACCÉDÉES VIA GBIF

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)
2. On doit accepter l'accord d'utilisation des données - <https://www.gbif.org/terms/data-user>
3. On doit documenter la façon dont vous traitez vos données
4. On doit citer correctement les données que vous utilisez
  - Lignes directrices - <https://www.gbif.org/citation-guidelines>
  - DOI des jeux de données dérivés - <https://www.gbif.org/derived-dataset/about>



# RÈGLES D'OR DE L'UTILISATION DES DONNÉES ACCÉDÉES VIA GBIF

1. On doit avoir un compte sur [www.gbif.org](http://www.gbif.org)
2. On doit accepter l'accord d'utilisation des données - <https://www.gbif.org/terms/data-user>
3. On doit documenter la façon dont vous traitez vos données
4. On doit citer correctement les données que vous utilisez
5. On doit déposer les données utilisées dans un référentiel public comme p.ex. Zenodo



	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
<a href="#">↓ SIMPLE</a>	X	✓	X	✓ (if available)	Tab-delimited CSV <a href="#">?</a>	<b>1 MB</b> (167 KB zipped for download)
<a href="#">↓ DARWIN CORE ARCHIVE</a>	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV <a href="#">?</a>	<b>3 MB</b> (423 KB zipped for download)
<a href="#">↓ SPECIES LIST</a>	X	✓	X	X	Tab-delimited CSV <a href="#">?</a>	

## TÉLÉCHARGEMENTS DE DONNÉES

Les données peuvent être téléchargées en trois formats

**Simple** : CSV délimité par des tabulations. Contient uniquement les données après interprétation du GBIF. Pas de multimédia inclus. Plus d'informations sur CSVArchive

**Darwin Core** : L'archive Darwin Core (DwC-A) contient à la fois les données originales fournies par l'éditeur et l'interprétation du GBIF. Liens (mais pas de fichiers) vers multimédia inclus. Plus d'informations sur DwC-A

**Liste des espèces** : CSV délimité par des tabulations avec la liste distincte des noms dans le résultat de la recherche.

# Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

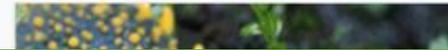
ABOUT GBIF DENMARK

Occurrence records  
1,676,825,999

Datasets  
57,757

Publishing institutions  
1,665

Peer-reviewed papers using data  
5,703



APERÇU DU SITE WEB

[www.gbif.org](http://www.gbif.org)

# Exercice 1: Naviguer [www.gbif.org](http://www.gbif.org)

Quel est le nombre total d'occurrences pour la Province du Haut Katanga en République Démocratique du Congo ?

Parmi ces occurrences, combien y a-t-il de la famille des abeilles Apidae dans la province ?

Combien d'occurrences ont été enregistrés à plus de 900 m d'altitude ?

Quel est le nombre d'entre eux permettant toute utilisation avec une attribution appropriée ?

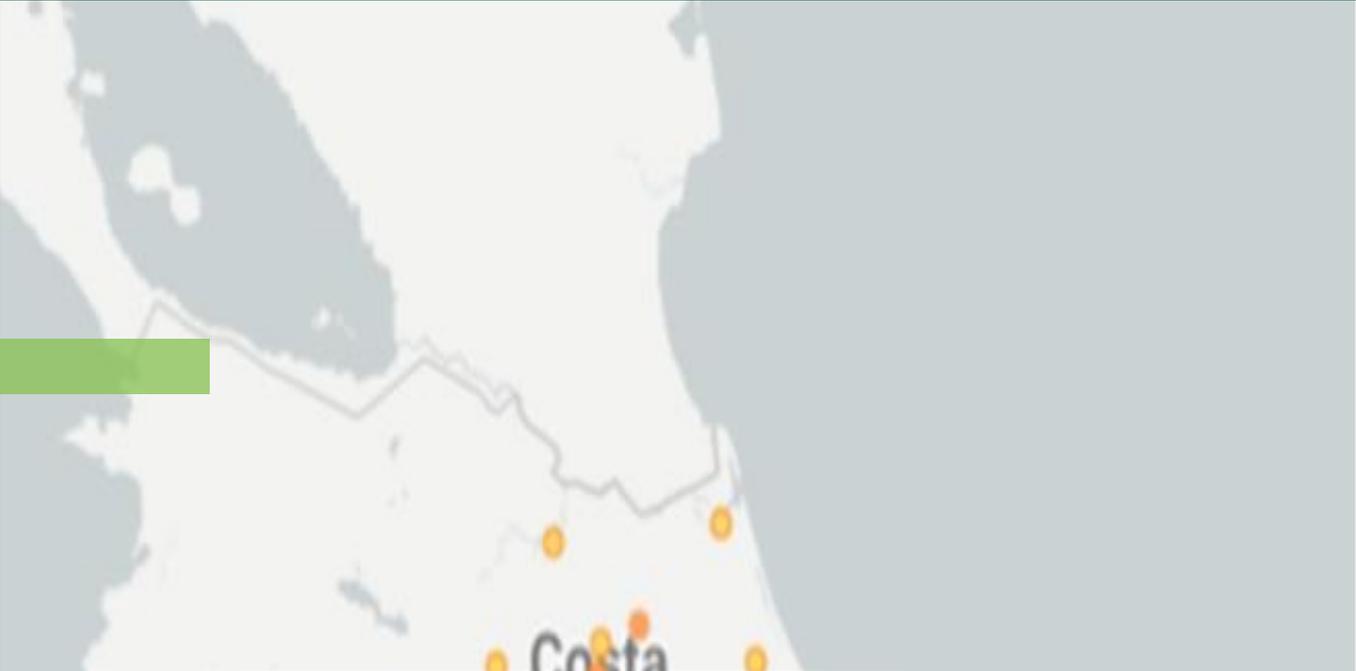
# Exercice 1: Naviguer [www.gbif.org](http://www.gbif.org)

Quel est le nombre total d'occurrences pour la Province du Haut Katanga en République Démocratique du Congo ? (45,165 ou 44, 675 - 23 Nov 2021)

Parmi ces occurrences, combien y a-t-il de la famille des abeilles Apidae dans la province ? (93 - 23 Nov 2021)

Combien d'occurrences ont été enregistrés à plus de 900 m d'altitude ? (44 - 23 Nov 2021)

Quel est le nombre d'entre eux permettant toute utilisation avec une attribution appropriée ? (44 – 23 Nov 2021)



# Problèmes courants de qualité des données

John Waller | Data Analyst



**GBIF**

Global Biodiversity  
Information Facility



# Post-traitement des téléchargements GBIF

Les données que vous téléchargez ne seront pas toujours suffisantes pour ce que vous voulez en faire. Il y a certaines choses que vous devez savoir...

- Enregistrements d'absence
- Fossiles
- Spécimens vivants
- Des zoos ou des jardins botaniques
- Lieu incertain
- Centroides des pays
- Dans l'océan
- Jeux de données quadrillées...

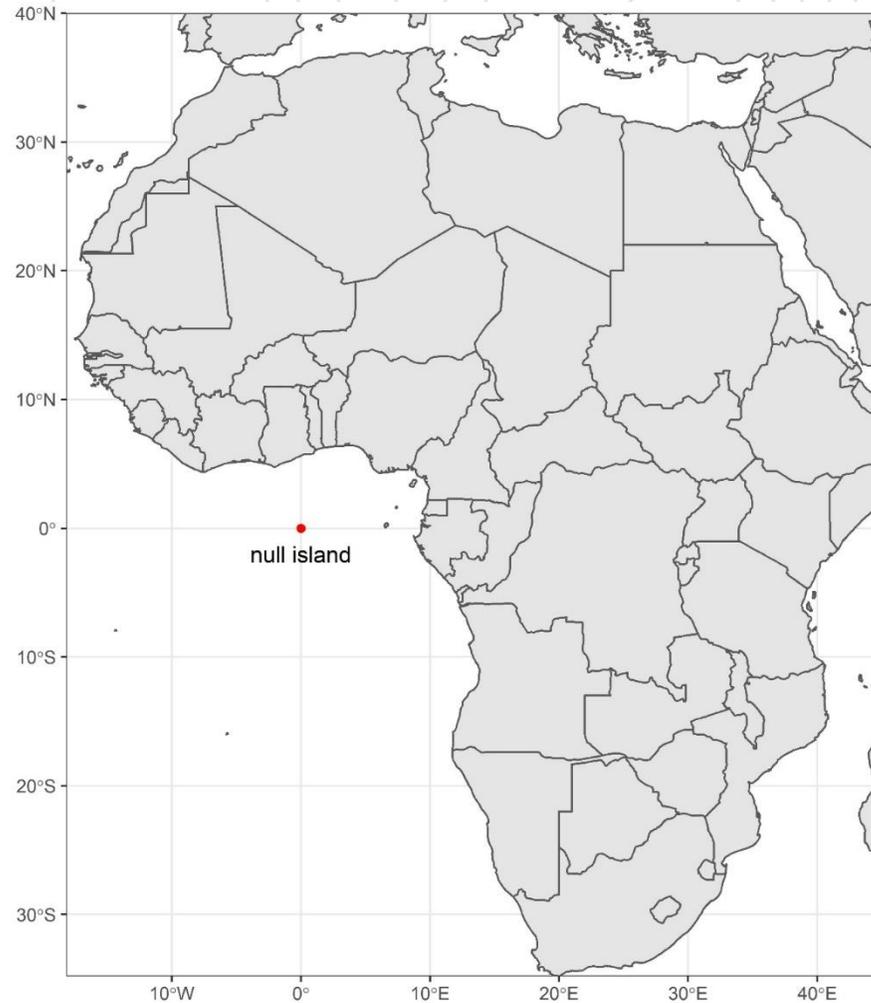
# Issues géospatiaux par défaut

Le GBIF supprime par défaut les issues géospatiaux courants si vous choisissez d'avoir des données avec un lieu.

Les éléments suivants seront supprimés :

- **Coordonnées zéro** : Les coordonnées sont exactement (0,0). île nulle
- **Non-concordance des coordonnées du pays** : les coordonnées se trouvent en dehors du polygone du pays donné.
- **Coordonnée invalide** : GBIF n'est pas en mesure d'interpréter les coordonnées.
- **Coordonnée hors limite** : Les coordonnées sont en dehors des limites pour les valeurs décimales lat/lon ((-90,90), (-180,180)).

# GBIF supprime les coordonnées zéro (0,0) « l'île nulle »

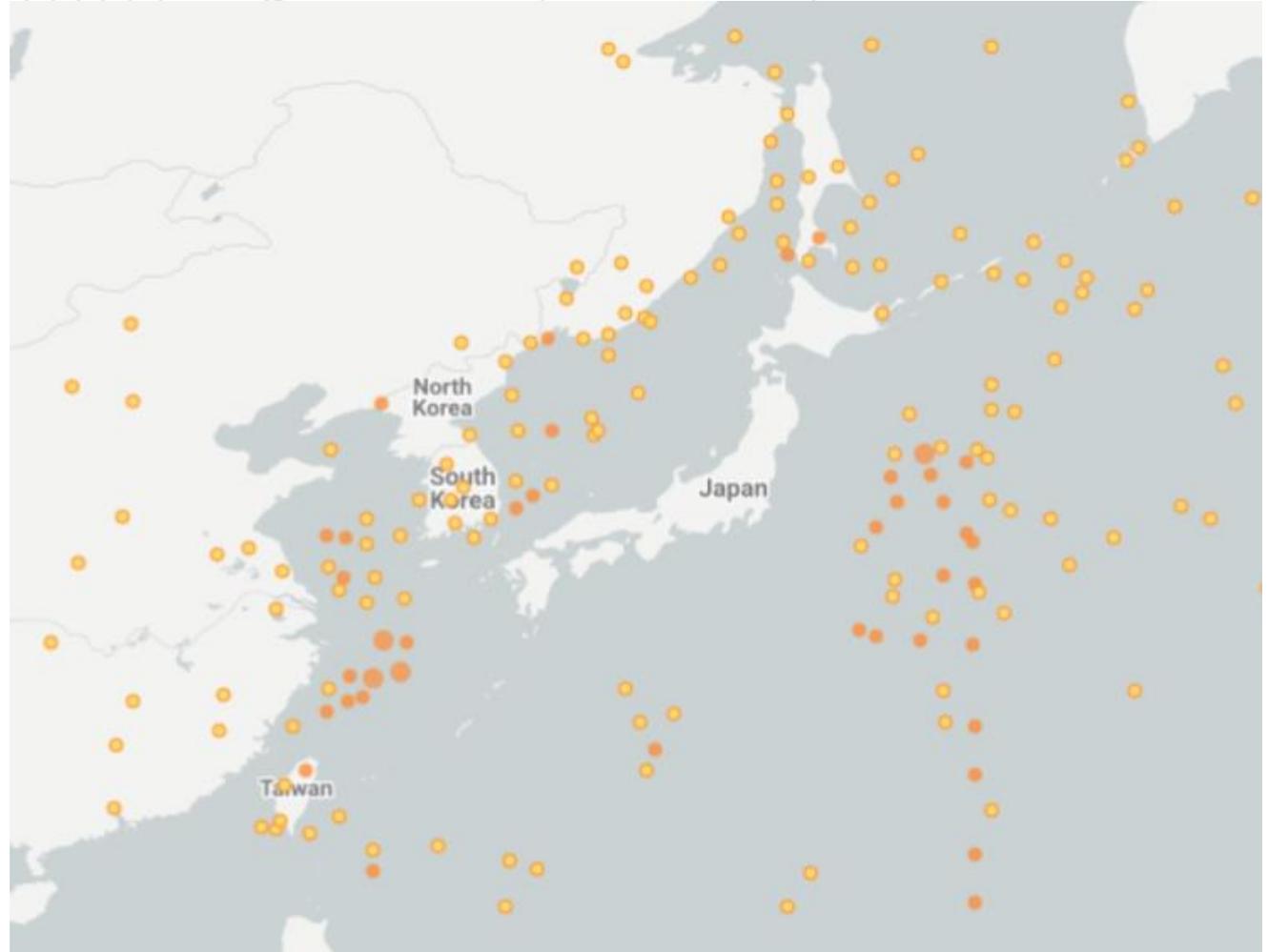


[https://www.gbif.org/occurrence/map?issue=ZERO\\_COORDINATE](https://www.gbif.org/occurrence/map?issue=ZERO_COORDINATE)

# Non-concordance des coordonnées du pays

GBIF supprime les enregistrements qui ne correspondent pas à leur countrycode.

Tous ces enregistrements prétendent être situés au Japon.



[https://www.gbif.org/occurrence/search?issue=COUNTRY\\_COORDINATE\\_MISMA](https://www.gbif.org/occurrence/search?issue=COUNTRY_COORDINATE_MISMA)

TCH

# Enregistrements d'absence

Parfois, les éditeurs de données incluront des enregistrements d'absence (où ils valident qu'une espèce n'est pas présente). Beaucoup d'utilisateurs ne veulent pas de ces enregistrements.

```
gbif_download %>%  
  filter(occurrenceStatus == "PRESENT")
```

[https://www.gbif.org/occurrence/search?occurrence\\_status=present](https://www.gbif.org/occurrence/search?occurrence_status=present)

# Fossiles et spécimens vivants

GBIF a des fossiles et des spécimens vivants (généralement une plante dans un jardin botanique ou parfois et un animal dans un zoo).

```
gbif_download %>%  
  filter(!basisOfRecord %in%  
  c("FOSSIL_SPECIMEN", "LIVING_SPECIMEN"))
```

# establishmentMeans

**dwc:establishmentMeans** : Le processus par lequel le ou les individus biologiques représentés dans l'enregistrement se sont établis à la localité.

```
gbif_download %>%  
  filter(!establishmentMeans %in% c("MANAGED",  
  "INTRODUCED", "INVASIVE", "NATURALISED"))
```

Malheureusement pas utilisé très souvent.

<https://terms.tdwg.org/wiki/dwc:establishmentMeans>

# Anciens enregistrements

Le GBIF possède de nombreux enregistrements de musée qui peuvent être plus anciens que ce qui est souhaité pour certaines études.

```
gbif_download %>%  
filter(year >= 1900)
```

[https://www.gbif.org/occurrence/search?year=1000,1650&occurrence\\_status=present](https://www.gbif.org/occurrence/search?year=1000,1650&occurrence_status=present)

# Exemple d'un lieu incertain

**Species:** *Lophodytes cucullatus* (Linnaeus, 1758)

**Location:** United States of America

**Basis of record:** Human observation

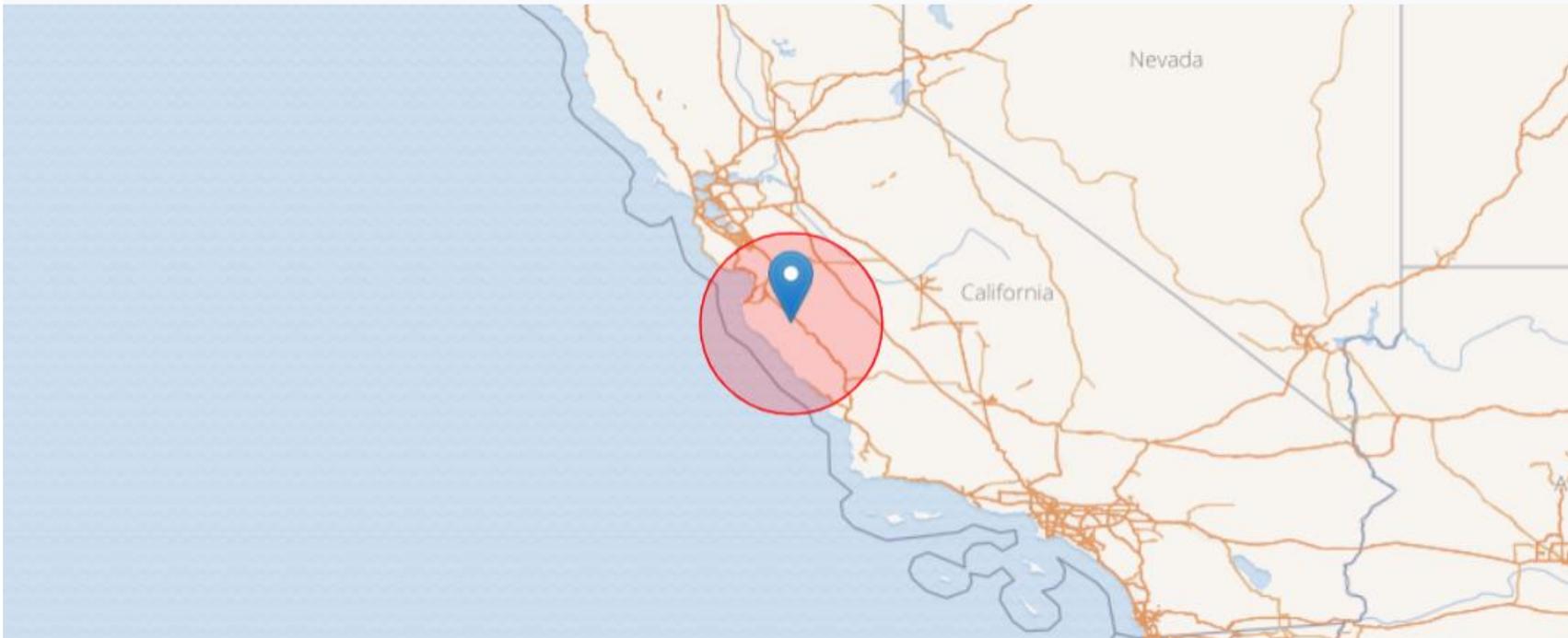


**Dataset:** iNaturalist Research-grade Observations

**Publisher:** iNaturalist.org

**Reference:** <https://www.inaturalist.org/observations/67427035>

**Issues:** Institution match none Collection match none



<https://www.gbif.org/occurrence/3017942707>

# Lieu incertain

Souvent, vous voudrez être sûr que les coordonnées donnent un certain lieu et ne sont pas vraiment à des milliers de kilomètres de l'endroit où l'organisme a été observé ou collecté.

```
gbif_download %>%  
filter(coordinatePrecision > 0.01 |  
is.na(coordinatePrecision)) %>%  
filter(coordinateUncertaintyInMeters < 10000 |  
is.na(coordinateUncertaintyInMeters))
```

Je recommande de ne pas filtrer les valeurs manquantes, car la valeur n'est souvent pas renseignée par les éditeurs s'ils pensent que l'occurrence est assez certaine (à partir d'un GPS).

# Mauvaises valeurs par défaut pour l'incertitude des coordonnées

```
gbif_download %>%  
  filter(!coordinateUncertaintyInMeters %in%  
  c(301, 3036, 999, 9999))
```

Il existe quelques « fausses » valeurs pour l'incertitude des coordonnées dont vous devez être conscient. Ces valeurs sont des erreurs produites par le logiciel de géocodage et ne représentent pas de vraies valeurs d'incertitude. Dans le cas de 301, l'incertitude est souvent bien supérieure à 301 et représente en fait un centroïde de pays.

# Points le long de l'équateur ou du premier méridien

Certains éditeurs considèrent que zéro et NULL sont équivalents, les champs latitude et longitude vides finissent par être tracés le long de ces deux lignes.

```
gbif_download %>%  
filter(!decimalLatitude == 0 |  
!decimalLongitude == 0)
```

<https://www.gbif.org/> (look at main map for patterns)

# Géocodage rétrospectif

Le géocodage rétrospectif est le processus par lequel les valeurs lat-lon sont données aux enregistrements plus anciens qui ne contiennent que des descriptions textuelles sur la localité.

Les informations sur la localité ne sont parfois qu'un pays, une ville ou une description textuelle comme

« 10 milles au sud-ouest de la route principale, Austin TX ».

Souvent, les enregistrements des musées (spécimens conservés) auront été géocodés rétrospectivement.

# Centroides de pays

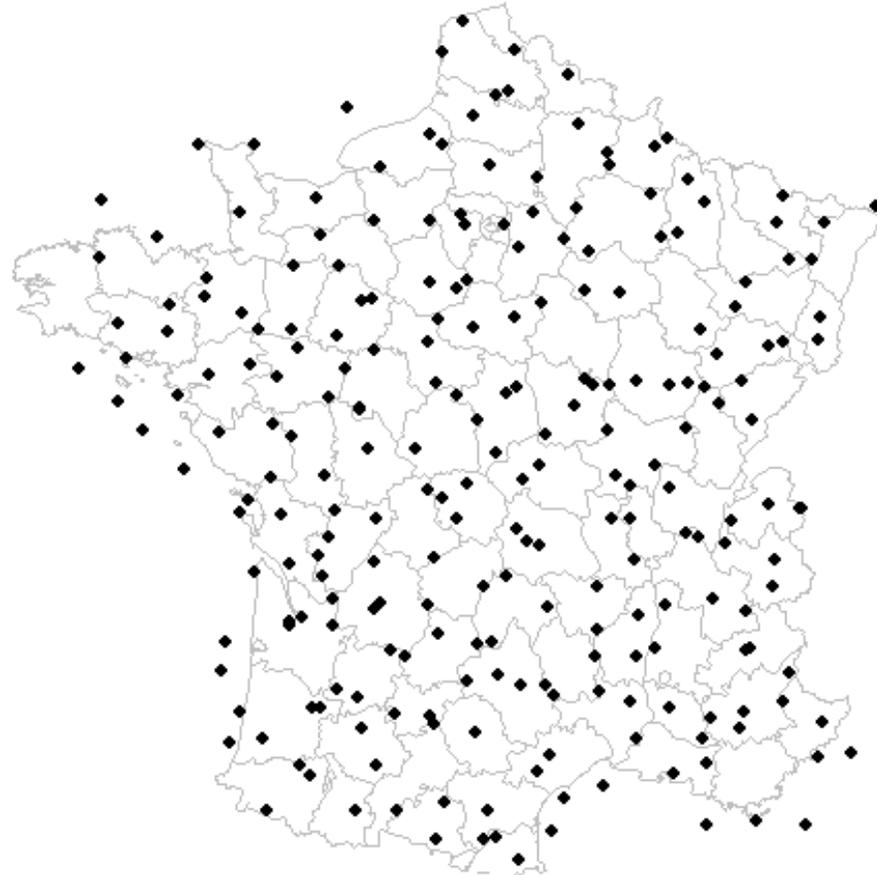


# Supprimer les doublons

Pour votre application, il peut être important de supprimer les enregistrements en double.

```
gbif_download %>%  
distinct(decimalLongitude, decimalLatitude, speci  
esKey, datasetKey, .keep_all = TRUE)
```

# Jeux de données quadrillées



<https://www.gbif.org/dataset/c779b049-28f3-4daf-bbf4-0a40830819b6>

# Filtrage de jeux de données quadrillées

La plupart des éditeurs d'ensembles de données maillées remplissent en fait l'une des colonnes suivantes :

- coordinateuncertaintyinmeters
- coordinateprecision
- footprintwkt (only in dwca downloads)

Le filtrage par ces colonnes peut donc être un bon moyen de supprimer des ensembles de données en grille.

Le GBIF dispose d'une [API expérimentale](#) pour identifier les ensembles de données qui présentent un certain degré de quadrillage. Vous pouvez lire plus [ici](#).

# Exercice 2 : Filtrage des données pour une meilleure qualité des données

À l'aide de [www.gbif.org](http://www.gbif.org), filtrez les données de *Calopteryx splendens* de la manière suivante :

- Filtrer pour les enregistrements avec une incertitude de coordonnées entre 0 et 10 000 m
- Filtrer les enregistrements entre 1955 et 2017
- Exclure les enregistrements d'occurrence où la moyenne d'établissement est indiquée comme gérée, introduite ou envahissante

Combien de disques aviez-vous au départ ?

Combien en avez-vous après filtrage ?

Comment avez-vous trouvé le taxon ?

Quelles sont les limites des filtres ?

# Occurrence API

<http://api.gbif.org/v1/>

SUMMARY REGISTRY SPECIES OCCURRENCE MAPS NEWS LITERATURE

This API works against the GBIF Occurrence Store, which handles occurrence records and makes them available through the web service and download files. In addition we also provide a Map API that offers spatial services.

Internally we use a Java web service client for the consumption of these HTTP-based, RESTful web services.

## Occurrences

This API provides services related to the retrieval of single occurrence records.

Resource URL	Method	Response	Description
--------------	--------	----------	-------------

## L'API GBIF

Andrew Rodrigues | Programme Officer



Please be aware that the following parameters are in an experimental phase and its definition could change in the future: q, facet, facetOffset, facetLimit, facetMincount and facetMultiSelect

Resource URL	Method	Response	Description	Paging	Parameters
/occurrence/search	GET	Occurrence	Full search across all occurrences. Results are ordered by relevance.	true	q, basicOfRecord, catalogNumber, classifi...

# Qu'est-ce que l'API GBIF ?

L'interface de programmation d'application (API) GBIF permet aux utilisateurs d'accéder aux bases de données GBIF de manière sécurisée

Habituellement, la principale raison pour laquelle vous voudriez utiliser une API est que vous voulez qu'un logiciel interagisse d'une manière ou d'une autre avec GBIF

L'API GBIF est accessible via :

Un navigateur Web en visitant une URL, par ex. [https://api.gbif.org/v1/species/match?name=Passer domesticus](https://api.gbif.org/v1/species/match?name=Passer%20domesticus)

Ou en utilisant un programme en ligne de commande appelé curl que vous devez installer

GBIF a [quelques groupes/espaces de noms d'API](#) :

**API de registre** - rend tous les ensembles de données, installations, organisations, nœuds et réseaux enregistrés détectables.

**API d'espèces** - fonctionne avec les données conservées dans la banque de listes de checklists GBIF qui indexe taxonomiquement tous les jeux de données de checklist enregistrés dans le réseau GBIF.

**API d'occurrence** - fonctionne avec le GBIF Occurrence Store, qui gère les enregistrements d'occurrences et les rend disponibles via le service Web et les fichiers de téléchargement.

**API cartographique** - service de tuile de carte Web permettant de visualiser facilement le contenu GBIF sur des cartes interactives et de superposer le contenu d'autres sources.

**API de littérature** - recherche de littérature indexée par GBIF, y compris des articles évalués par des pairs, citant des ensembles de données et des téléchargements GBIF.

# Qu'est-ce que l'API GBIF ?

Le modèle de base d'un appel d'API ::

- **URL de base** : ce sera toujours **https://api.gbif.org/v1/**
- **api** : c'est le groupe/espace de noms d'API GBIF que vous souhaitez interroger.
- **fonction** : la fonctionnalité que vous souhaitez utiliser
- **paramètre** : les paramètres de votre appel API. Une « ? » est parfois utilisé
- **requête** : la requête que vous remplissez. Parfois, ce sera du texte libre et parfois un argument prédéfini.

Exemple

<https://api.gbif.org/v1/species/match?name=Passer domesticus>

# Exercice 3 - Recherche de taxonkeys GBIF

- Les taxonkeys sont attribués au niveau de l'espèce, de la famille du genre, de l'ordre, du phylum et du royaume.
- Des identifiants uniques sont délivrés aux noms acceptés avec des synonymes de ces noms acceptés délivrés avec le même identifiant.
- Permettre à l'utilisateur de s'assurer qu'il collecte toutes les données dont il a besoin
- Facilitez également les téléchargements de plusieurs espèces
- Les clés de taxon peuvent être trouvées via :
  - [www.gbif.org](http://www.gbif.org)
  - API d'espèces
  - Outil de Species Matching - <https://www.gbif.org/tools/species-lookup>

<https://api.gbif.org/v1/species/match?name=Calopteryx%20splendens>

## Exercice 3 - Recherche de taxonkeys GBIF

À l'aide de l'API Species, recherchez les taxonkeys GBIF pour ces noms scientifiques :

- *Lepus saxatilis* F.Cuvier, 1823
- Aves
- Magnoliophytes
- *Aegithalos caudatus* (Linnaeus, 1758)

Quel est le statut taxonomique de chacun ?

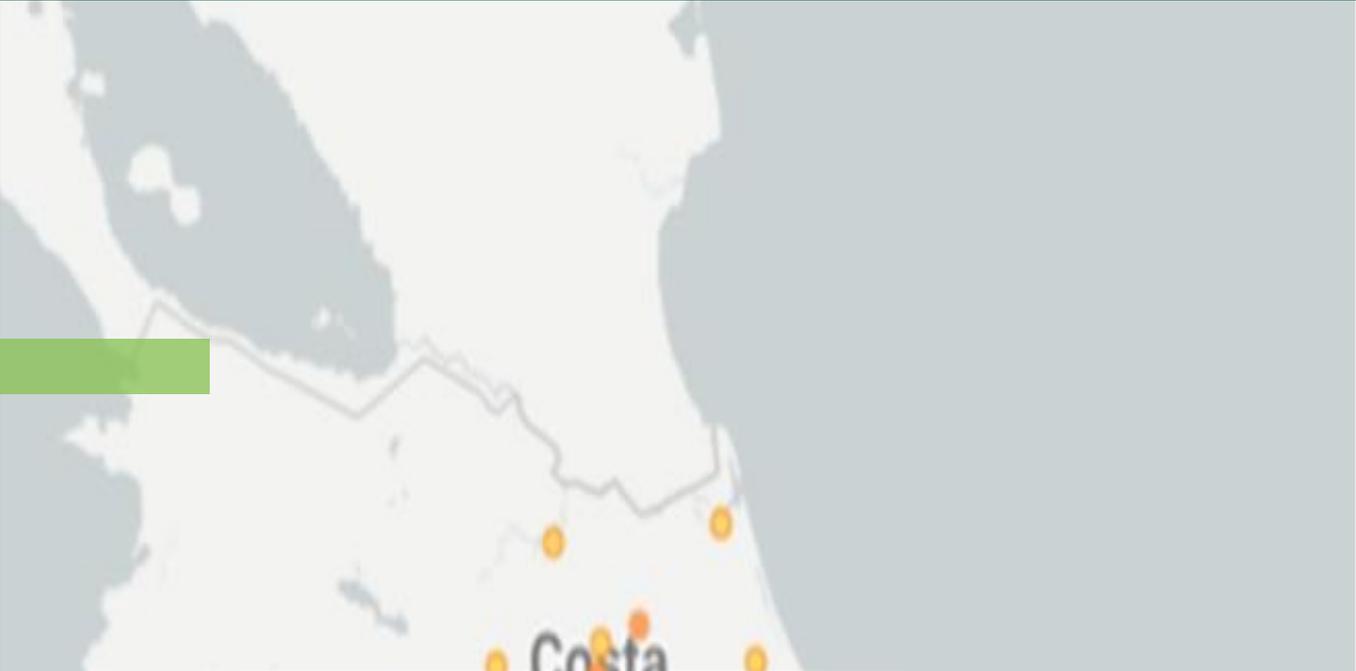
# Exercice 3 - Recherche de taxonkeys GBIF

À l'aide de l'API Species, recherchez les taxonkeys GBIF pour ces noms scientifiques :

- *Lepus saxatilis* F.Cuvier, 1823 = 2436775 (ACCEPTED)
- Aves = 212 (ACCEPTED) Remarque : il ne s'agit pas d'une espèce, donc si vous voulez des occurrences pour un groupe entier, vous n'avez besoin que d'un taxonkey et non d'une liste de toutes les espèces du groupe..
- Magnoliophyta = 49 (SYNONYM). Remarque : si vous avez utilisé l'API directement, GBIF vous donnera la clé d'utilisation acceptée : 7707728, qui est Tracheophyta. C'est le nom accepté de ce groupe. Si vous décidez de n'utiliser que l'ancien nom, vous passeriez à côté de millions d'occurrences, alors soyez prudent.
- *Aegithalos caudatus* (Linnaeus, 1758) = 2495000 (DOUBTFUL) Remarque : il s'agit d'un cas intéressant car ce nom « douteux » a des millions d'occurrences qui lui sont liées. Il y a une histoire taxonomique apparemment intéressante derrière cette affaire...

Quel est le statut taxonomique de chacun ?

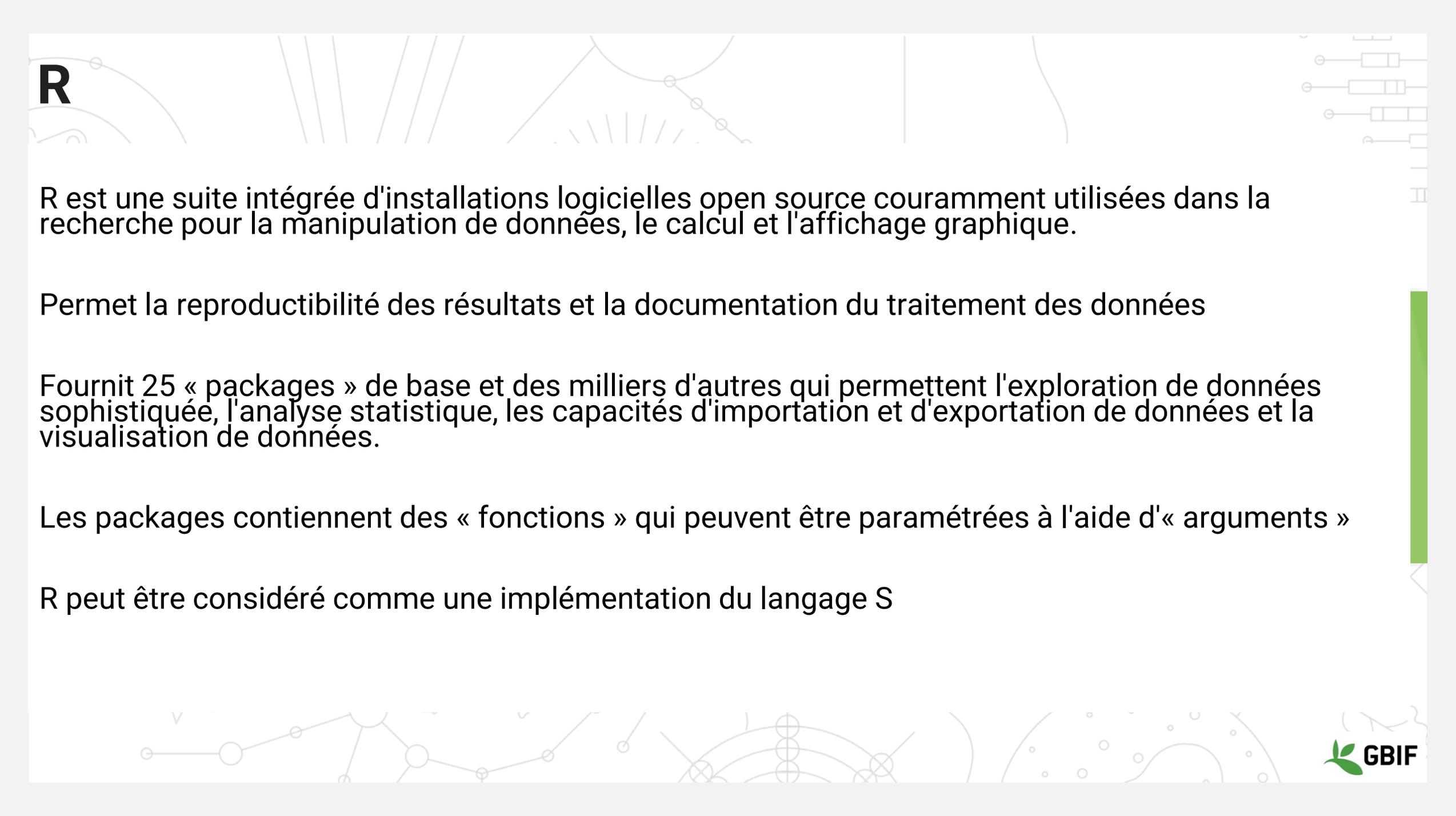
<https://api.gbif.org/v1/species/match?name=Calopteryx%20splendens>



**R et rgbif**

**John Waller | Data Analyst**





# R

R est une suite intégrée d'installations logicielles open source couramment utilisées dans la recherche pour la manipulation de données, le calcul et l'affichage graphique.

Permet la reproductibilité des résultats et la documentation du traitement des données

Fournit 25 « packages » de base et des milliers d'autres qui permettent l'exploration de données sophistiquée, l'analyse statistique, les capacités d'importation et d'exportation de données et la visualisation de données.

Les packages contiennent des « fonctions » qui peuvent être paramétrées à l'aide d'« arguments »

R peut être considéré comme une implémentation du langage S

# Les bases de R

Possède une interface de ligne de commande bien qu'une interface utilisateur graphique, par ex. Rstudio

L'invite par défaut est '>'

Les commandes sont séparées soit par un point-virgule (;), soit par une nouvelle ligne

Une syntaxe simple, sensible à la casse et permettant tous les symboles alphanumériques est autorisée (et dans certains pays, les lettres accentuées) plus '.' et '\_'

Les noms doivent comporter « . » ou une lettre, et s'ils commencent par « . », le deuxième caractère ne doit pas être un chiffre. Les noms sont effectivement d'une longueur illimitée.

# Syntaxe R

Les commandes sont séparées soit par un point-virgule (;), soit par une nouvelle ligne

Les commentaires peuvent être placés presque n'importe où, en commençant par un dièse ("#"), tout jusqu'à la fin de la ligne est un commentaire.

Si une commande n'est pas terminée à la fin d'une ligne, R donnera une invite différente, par défaut

Les entités que R crée et manipule sont appelées objets. Au cours d'une session R, les objets sont créés et stockés par nom. La collection d'objets actuellement stockés est appelée l'espace de travail.

# Syntaxe R

La structure de base d'une commande est la fonction (arguments). Les parenthèses sont utilisées pour entourer les arguments que vous utiliserez.

Les opérateurs sont des symboles qui indiquent au compilateur d'effectuer des manipulations mathématiques ou logiques spécifiques, par ex. +, -, \*, /

Les variables peuvent se voir attribuer des valeurs à l'aide de l'opérateur gauche, droite et égal à.

Une variable nous fournit un stockage nommé que nos programmes peuvent manipuler.

# Démo de RStudio

# Exercice 4: Configuration de R

En utilisant RStudio, vous allez ouvrir votre document Rmarkdown et installer les packages dont vous avez besoin.

```
#install the packages you will need and read in their libraries
install.packages("tidyverse")
install.packages("rgbif")
install.packages("taxize")
install.packages("CoordinateCleaner")
library(rgbif)
library(dplyr)
library(readr)
library(taxize)
library(CoordinateCleaner)
```

Vous allez installer les packages suivants qui accèdent à la base de données GBIF via l'API :

- **tidyverse** - tidyverse inclut les packages que vous êtes susceptible d'utiliser dans les analyses de données quotidiennes
- **rgbif** - rechercher et récupérer des données du Global Biodiversity Information Facility (GBIF)
- **taxize**- interagit avec une suite d'API Web pour les tâches taxonomiques
- **CoordinateCleaner** - Nettoyage automatisé des enregistrements d'occurrences des collections biologiques

Vous utilisez **install.packages()** et **update.packages()** pour installer et mettre à jour les packages

Vous utilisez la fonction **library()** pour charger un package particulier dans votre session R

# Qu'est-ce que rgbif ?

rgbif vous donne accès aux données du GBIF via leur REST API

La documentation peut être trouvée ici: <https://www.rdocumentation.org/packages/rgbif/versions/1.2.0>

Peut exécuter une gamme de fonctions en utilisant les différentes API mises à la disposition des utilisateurs, y compris

- Effectuer des recherches dans le registre du GBIF
- Récupérer les cartes des données d'occurrence
- Récupérer des informations taxonomiques à partir des données véhiculées par le GBIF
- Télécharger les données médiées par le GBIF
- Rechercher des données d'occurrence

<https://docs.ropensci.org/rgbif/reference/index.html>

# Fonctions utiles de base

`occ_download()` - lance une demande de téléchargement pour les données d'occurrence GBIF.

`pred ()` - fonction de filtrage qui vous permet uniquement de spécifier une valeur que vous souhaitez filtrer

`pred_in()`- fonction de filtrage qui vous permet uniquement de spécifier plusieurs valeurs que vous souhaitez filtrer

`occ_download_get()` - apporte votre téléchargement GBIF dans votre répertoire de travail

`occ_download_import()` - charge vos données dans votre console R

`occ_download_meta()` - récupère les métadonnées de téléchargement ; utile pour récupérer les DOI

N'oubliez pas que la fonction d'aide intégrée de R est utile si vous ne savez pas quels arguments vous devez utiliser pour des fonctions spécifiques

```
> help(solve)
```

ou

```
> ? solve
```

# Exercice 5 : Téléchargement d'un jeu de données à l'aide de rgbif

Téléchargez un jeu de données (CSV simple) à l'aide de rgbif qui possède les propriétés suivantes

- Le taxon est *Lepus saxatilis*
- Trouvé en Afrique du Sud (ZA)
- Un spécimen préservé ou une observation humaine
- Qui a des coordonnées de latitude et de longitude
- N'a pas de problèmes géospatiaux communs

# Exercice 5 : Téléchargement d'un jeu de données à l'aide de rgbif

```
library(rgbif)
```

```
user=""  
pwd=""  
email=""
```

```
occ_download(  
  pred("taxonKey", ?),  
  pred_in("basisOfRecord", c('PRESERVED_SPECIMEN', 'HUMAN_OBSERVATION')),  
  pred("country", "ZA"),  
  pred("hasCoordinate", TRUE),  
  pred("hasGeospatialIssue", FALSE),  
  format = "SIMPLE_CSV",  
  user=user, pwd=pwd, email=email  
)
```

[link to taxon](#)

# 1ère étape - connectez-vous au GBIF

```
user=""  
pwd=""  
email=""
```

# 2ème étape - Utilisez la fonction `occ_download`

```
occ_download(  
  pred("taxonKey", 2436775), #Taxon TBD  
  pred_in("basisOfRecord", c('PRESERVED_SPECIMEN', 'HUMAN_OBSERVATION')),  
  pred("country", "ZA"),  
  pred("hasCoordinate", TRUE),  
  pred("hasGeospatialIssue", FALSE),  
  format = "SIMPLE_CSV",  
  user=user, pwd=pwd, email=email  
)
```

Ce code définit vos paramètres de filtrage et de téléchargement

- Taxon key
- Basis of record
- Country
- hasCoordinate
- hasGeospatialIssue
- Format

# 3ème étape - Importation de votre téléchargement dans R

```
#read your data download into R and obtain a DOI for your dataset
data_download <- occ_download_get(gbif_download_key, overwrite= TRUE)%>%
occ_download_import()
View(data_download)
res <- occ_download_meta(gbif_download_key)
gbif_citation(res)
```

%>% est un opérateur de type pipe avec lequel vous pouvez transférer une valeur dans le prochain appel de fonction du package *magrittr* et permet l'imbrication de fonctions dans votre code

View(data\_download) - vous permettra de voir votre ensemble de données dans R

res <- créera une variable dans votre espace de travail appelée « res » qui contiendra les métadonnées de votre téléchargement

gbif\_citation()- extraira les données DOI des métadonnées

# Exercice 6 : Téléchargement d'une longue liste d'espèces

Utilisez les taxon keys que vous avez obtenues à partir de l'exercice 3 pour télécharger les données d'occurrence d'une longue liste d'espèces, par ex.

```
gbif taxon_keys =  
c(3189834,  
  3189801,  
  2876099,  
  3023817,  
  3033893,  
  2875008,  
  2875770,  
  3082226,  
  2875967,  
  3025563,  
  2984510,  
  2888580)  
  
occ_download(  
  pred_in("taxonKey", gbif taxon_keys),  
  pred("hasCoordinate", TRUE),  
  pred("hasGeospatialIssue", FALSE),  
  format = "SIMPLE CSV",  
  user=user, pwd=pwd, email=email  
)
```